Techniques for Extracting High-resolution Still Images from Video Sequences

Robert Whatmough

DSTO-TR-1247

20020509 043

Techniques for Extracting High-resolution Still Images from Video Sequences

Robert Whatmough

Surveillance Systems Division
Electronics and Surveillance Research Laboratory

DSTO-TR-1247

ABSTRACT

This Report describes techniques for extracting a single frame or part of a frame from a video image sequence and combining information from other frames to enhance the resolution of the result. The result depends on accurate alignment of the frames and construction of a detailed image consistent with the coarse data in the frames. Techniques for the alignment and the construction steps, and theoretical limitations on the final resolution, are considered.

RELEASE LIMITATION

Approved for public release

DEPARTMENT OF DEFENCE
DEFENCE SCIENCE & TECHNOLOGY ORGANISATION

DSTO

AQ F02-08-1538

Techniques for Extracting High-resolution Still Images from Video Sequences

Executive Summary

Video image sequences, including those produced by popular hand-held cameras, are a useful source of military and forensic information. They often suffer from poor resolution, made worse by poor focus, movement of the camera or objects of interest, and limitations of the recording medium. They are also affected by electronic noise in the camera or playback equipment, tape noise, and distortion of moving objects caused by the scanning process, especially by the commonly used technique of interlacing.

A sequence usually shows an object of interest in many frames. If the object remains still during part of the sequence, there is a redundancy of information which might be used to reduce noise, perhaps by averaging. If the view changes, through motion of the object or camera or a change in zoom, additional information is available which might make it possible to extract finer details than can be resolved by the pixels in a single frame.

In the most favourable case, camera positions can be set up so that the sequence is equivalent to a single image with smaller pixels but with blur, and the blur can be removed by standard image restoration methods. There is a limit to the reduction of pixel size, beyond which not all the details of the scene can be recovered correctly. This limit can be as low as a factor of 2, but for real cameras may be 3 or more, depending on internal details and perhaps the orientation of features of interest.

Usually the amount of movement between frames varies across the scene and must be measured from the image data. If a single object is of interest, and it moves slowly without change of distance or orientation, it is still relatively easy to estimate its movements accurately. Despite any irregularities of its motion, a method is then available to form a single image with improved resolution for that object. There are complications if the sequence came from an interlaced video system or if it is affected by some kinds of interference fringes. More rapid motion leads to motion blur, which needs treatment during the extraction process. In the worst case, motion must be estimated for different objects or even for different parts of one object. The resolution improvements can still be made, but they depend on the motion in the scene and on the quality of motion estimation. For uniform motion in one direction, the improvement may be only in that direction; if there is no motion, resolution cannot be improved at all.

If an object changes its distance from the camera, or the photographer varies the camera zoom or distance from the object during the sequence, further gains may be possible when a long enough sequence is available. In any case, the number of pixels in

all input frames combined must be at least the number required in the result, so a resolution gain of 3 requires at least 3x3 or 9 frames of input.

Tests with simulated image data showed that resolution improvements of 2 or 3 were quite possible so long as there were more pixels in the low resolution frames than were wanted in the high resolution output, and the registration was accurate within 0.1 of a pixel in the output.

This Report presents examples of high resolution extraction, and describes alternative methods of extraction (mostly rejected because of unreliable performance or lack of speed). It does not give detailed descriptions, but is a summary of what was done, what was learned and where further study can be done.

Authors

Robert Whatmough
Surveillance Systems Division

Robert Whatmough was employed as Cadet Defence Science at the Weapons Research Establishement, Department of Supply, later DSTO. He was awarded the degree of B Sc with honours in 1969.

Until 1985 he was Experimental Officer, Scientific Officer then Research Scientist in Computing Services Group. His work included mathematical data handling techniques, random number generation, printed circuit board design, simulation of batch computer operations, curve fitting and smoothing, time-critical computing, thermal modelling, computer graphics and display, remote sensing and image processing, and assistance of computer users with complex computing problems.

Since 1986 he has worked in various Divisions in fields related to image processing. These have included restoration, enhancement and classification of visual, multispectral and synthetic aperture radar images, geographic information systems, prediction of oblique aerial and navigation radar images, matching models to objects in images, shape inference from perspective distortion, registration and mosaicking of aerial images, and enhancement of video sequences.

Contents

1. Introduction

Video image sequences, including those produced by popular hand-held cameras, are a useful source of military and forensic information. They often suffer from poor resolution, made worse by poor focus, movement of the camera or objects of interest, and limitations of the recording medium. They are also affected by electronic noise in the camera or playback equipment, tape noise, and geometric distortion caused by the scanning process, especially by the commonly used technique of interlacing.

A sequence usually shows an object of interest in many frames. If the object remains still during part of the sequence, there is a redundancy of information which might be used to reduce noise, perhaps by averaging. If the view changes, through motion of the object or camera or a change in zoom, additional information is available which might make it possible to extract finer details than can be resolved by the pixels in a single frame.

In this Report, Section 2 reviews extensive work done in this area since 1988. Section 3 describes the video resolution enhancement problem and considers the resolution improvement when object motion is ideal. Section 4 considers the more realistic case where the motion is more general and of unknown type, and interlacing may be in use. Section 5 describes a resolution enhancement method developed for a single, slowly moving object. Section 6 describes alternative approaches for more general cases, including motion blur removal. Appendix A describes alternative algorithms considered for resolution improvement but found to be inadequate for the purpose.

2. Previous work

Many studies have been made of the improvement of resolution in video sequences by alignment and combination of frames. These differ in the number of assumptions made about the motion between frames and about details of the imaging process, and in the methods used to combine frames. Some of the methods will be defined in detail elsewhere in this Report.

A crude method of resolution enhancement is to warp the frames back to a common coordinate system, average, then reduce blur. This has been used on separate satellite images [Albertz & Zelianeos, 1990] with a claimed resolution improvement by a factor of 2 if registration of frames is accurate to within 0.1 pixels.

A better approach for greater resolution improvement is to map each pixel in the video frames to a position in the high-resolution image, find a value for each high-resolution pixel from a weighted mean of the nearest few video pixels, then remove known optical blur by Wiener filtering. This gives fairly good results for a resolution gain of 4 [Alam & Bognar, 1997]. A variant of this approach, tried only for simple shifts between frames, involves associating each video pixel with the nearest high-resolution output

pixel, then finding one value for each output pixel by averaging where there is more than one association or interpolating when there is none [Gillette *et al*, 1995]. This variant was extended to correct for hardware sensitivity differences between pixels [Armstrong *et al*, 1999].

Relative warping of frames, focus blur and illumination changes between frames have been handled by interpreting the grey-level gradients in each frame as blurry edges, sharpening, then merging the edges using a median or maximum operation [Chiang & Boult, 1997a]. (Resolution was increased by a factor of 4 but results were not sharp.)

Back-projection (see Section A.1) has been used as the basis of iterative methods to construct an image from frames with different warping, starting from the average of re-aligned frames [Irani & Peleg, 1991]. It has been combined with tracking of different objects in a scene [Irani & Peleg, 1993]. It has also been applied to sequences with mechanical vibration, after automatic selection of the best subset of the frames to use [Stern *et al*, 1999].

Projection Onto Convex Sets (POCS, see Section A.5) was used to combine frames in an early study [Stark & Oskoui, 1989], allowing only rotation and translation between frames (in this case, tomographic images). It was later modified to allow for measurement noise [Tekalp *et al*, 1992], [Patti *et al*, 1997] and motion blur [Patti *et al*, 1994]. Other enhancements correctly handled object motions, motion and focus blur, and occlusions (areas hidden by moving objects in some frames only) [Eren *et al*, 1996]. Further visual improvements have been made by using more complex models of image formation and reducing the estimated blur across edges to avoid "ringing" effects [Patti & Altunbasak, 1998, 2001].

A much more efficient use of POCS, fitting a whole video frame at a time, can be used if all frames are related by affine transformations (i.e. shifting, rotation, scaling and shearing). It uses the fast fractional Fourier transform (FFRFT) method to apply the transformations and has been adapted to cope with illumination changes as well [Granrath & Lersch, 1998].

The Bayesian or Maximum *a Priori* (MAP) approach tries to construct the most probable high-resolution image of the scene, given a prior probability model of the formation of the scene and a conditional probability model of recording the image frames (including noise), in accordance with Bayes' theorem. The construction usually reduces in practice to minimum mean square error (MMSE) estimation with regularisation (Sections 5.2, 5.3). Cheeseman *et al* [1994] did this iteratively, allowing unknown registration parameters to be found as part of the minimisation. Hardie *et al* [1997a] allowed rotations and translations and tried gradient descent and conjugate gradient methods in MMSE estimation. Hardie *et al* [1997b] extended this approach to differently moving objects (with motion estimation included) and got fair results with resolution increase by factors of 4 to 5. Similar gains have been obtained with movements estimated first by optical flow methods [Hardie *et al*, 1997c].

A prior probability model of the scene, the Huber-Markov prior, allows sharp edges in the result when used in the MAP approach. This has been applied to interlaced images with differently moving objects [Schultz & Stevenson, 1996], using quadratic maximisation and gradients projected onto a subspace that forces input pixel values to be reproduced. (Noise-free input was assumed.) Motion blur and noise were later allowed with motion estimation included in the maximisation [Schultz & Stevenson, 1998], or derived from movement information encoded in MPEG sequences [Chen & Schultz, 1999], [Sale & Schultz, 1999].

Other prior probability models have been considered for images with sharp edges, and the best choice appears to depend on the image type and the time available for calculations [Lorette et al, 1997].

Alternatives to the MAP approach can be derived from information theory. These have been used to handle motion blur that varies across the scene [Tull & Katsaggelos, 1995], or unknown focus blur [Leung & Lane, 2000]. The case of uniform motion and motion blur and unknown focus blur has been handled by minimising a combination of mean square error and total variation (or mean absolute gradient) iteratively [Lai & Cui, 1999].

A mixture of MAP and POCS methods, applied iteratively, may be useful for variable blur and noise. [Elad & Feuer, 1997].

Exact algebraic construction of a higher-resolution image is possible in special cases, usually involving simple transformations to align frames. The Singular Value Distribution can be used for simple translations [Hildebrandt & Newsam, 1990]. Assumptions about band-limitedness have also helped [Ur & Gross, 1992], [Luca et al, 1997]. These methods can break down if two frames almost coincide (e.g. if motion is temporarily slow).

If a high-resolution version of one frame has been prepared from that frame and previous frames, it is possible to use it and the next frame to form a high resolution version of that. A higher-resolution sequence is produced, with poorer results for the first few frames. This has been attempted using the POCS approach [Avrin & Dinstein, 1997], the MMSE approach applied iteratively [Elad & Feuer, 1999a], and Kalman filtering [Elad & Feuer, 1999b].

3. Enhancement with Ideal Motion

3.1 Image Formation

A video camera generally comprises:

- An optical system, which forms an optical image of an external scene on an internal focal plane;

- A transducer, which forms an analogue or digital signal from the optical image, at the focal plane;

- Mechanisms to adjust the focus and zoom of the optical system and brightness and contrast levels of the signal, perhaps partly automatically; and

- A recorder to retain the signal for later use.

It is assumed that any analogue camera is used in conjunction with a video capture system that converts the analogue signal to digital form. Henceforth only digital video signals are considered.

The digital video signal is a sequence of frames, each of which approximately represents the image on the focal plane at an instant of time. A frame is a sequence of samples (or pixels), each of which is a single number representing a radiance, or several numbers (usually three) representing colour. Colour signals can be processed by an extension of the methods in this Report, but are not considered further.

Each sample in a frame relates to a particular location in the focal plane. The same locations are represented in the same order in each frame, and are generally arranged on a rectangular (and approximately square) grid. The samples in a digital camera can be measured simultaneously, but those from an analogue camera are necessarily measured in sequence, and may be staggered in time over most of the period from the starting time of one frame to that of the next. Staggered samples are taken in order along one horizontal line at a time and, in the simplest case, the lines are considered in order from top to bottom. However, it is common in analogue cameras to scan the lines in an interlaced fashion, first the even-numbered lines then the odd. (The top line is numbered zero.) It may even happen, in infra-red work, that lines are scanned in four sets. In these cases, the lines sampled in one pass from top to bottom are referred to as a "field" of the frame.

Some limitations of the image scanning process can now be described:

- The transducer cannot respond to instantaneous light levels in the optical image. Any sample taken must cover a length of time, and any movement of objects in the image will cause motion blur.

- The transducer responds to the average radiance over some small area. Even if the image is stationary, a sample does not respond to the radiance at a single point, nor does it respond to the whole rectangular area that is closer to the nominal sample location than to any other. The area actually sampled may lie between these extremes. (The ratio of the sampled area to the whole is known as the "fill factor".)

- If the sampling is sequential, and objects are moving in the image, distortion of objects will occur. Line-by-line scanning within a field will cause horizontal shearing or vertical expansion or contraction. Interlacing and motion together will cause the two fields to be misaligned, and ragged object edges to appear in the digital image.

The spacing between pixels is often the main limiting factor for image resolution. Motion blur and poor focus can further degrade the resolution. Interlacing together with motion can severely degrade the appearance of a single frame, but the effect of sequential scanning with motion can even be ignored if the motion is uniform and objects take many frames to cross the field of view.

3.2 Ideal motion

Suppose that the camera can be moved so that:

- Each frame shows the same scene shifted by a slightly different amount without any other change (other than gains or losses of scene strips at the edges)

- All movements are completed between scans

- The shifts are multiples of the same submultiples of the pixel spacing (eg, 0, 1/3 and 2/3 pixels horizontally, and similarly in the vertical direction)

If all combinations of horizontal and vertical offsets are used once each in different frames of a sequence, then the frames jointly contain samples of the same extent as before, but with a reduced and still regular spacing. If the samples are re-ordered, they form a single image in which the sample spacing is smaller, but the sample size (amount of focal plane covered by one sample) is unchanged. Now that the limitation is the size rather than the spacing, the image is a candidate for de-blurring.

The above approach has been considered as a method for improving the resolution of a digital still camera [Luca *et al*, 1997]. Each image would be recorded as a sequence of images mechanically displaced, then their pixels would be re-ordered and de-blurring applied. Hildebrandt & Newsam [1990] considered the case where there are four images, with known displacements only approximately 0 and ½ pixels each way, and a pixel size reduction by 2 each way is required.

3.3 Limitations of de-blurring

The blur to be removed from the combined image formed above is the effect of using a sample size greater than the sample spacing. It is roughly equivalent to applying a moving window mean to samples whose size matches the spacing. The effect of the blur is best considered in terms of the Fourier components of the image.

When a sample is taken of a single Fourier component, there will usually be some values of its frequency for which the result will be zero or very small. (For example, if the sample is equivalent to a 2x2 block of pixels, the frequency values are the ones for which the horizontal or vertical part is a non-zero multiple of one cycle per block, or half a cycle per pixel.) The smallest such value will be about one cycle in the width of the sample, the exact value depending on the orientation of the Fourier component and how the sensitivity of the sample falls off near its edge. If there is detail in the optical image that requires pixels smaller than half the sample width, then there will be Fourier components with frequencies higher than one cycle per sample width, and usually some components at that frequency which will be missing from all measurements and not recoverable by de-blurring. Any attempt to construct an image with pixels smaller than half the sample width will then produce a misleading result, because the presence of components with frequencies at and above one cycle per sample width is implied. (Even when the image dimensions make it impossible for any component to have exactly the frequency that will cause it to disappear, some components will come so close that they are represented mainly by noise in the input frames, and these must be suppressed by correct regularisation.)

In the simplest case, where the full area of each input image pixel is sampled, correct restoration cannot be expected if this area exceeds a 2x2 block of output pixels, so the resolution gain has a practical limit of 2. In the more likely case, where a smaller area is sampled, the output pixels can be made correspondingly smaller and the limit is greater than 2. Of course, there must be enough frames to cover the whole area to be restored with samples; the number of input pixels in all frames must be at least the number of output pixels in the higher resolution frame.

Only by changing the sample size (say by altering the focal length of the camera or changing the distance from camera to object) can the missing Fourier components be recovered; those actions are outside the scope of ideal motion.

Optical blur or motion blur can reduce the resolution limit. Typically a point source is projected through an out-of-focus lens system to a circular disk or polygon, depending on details of the aperture. If the spread is larger than the sample size, components will be destroyed at frequencies lower than the lowest destroyed by the sampling, and details implied by the use of pixels smaller than about half the blur size will not be restored. Changes to the focus or motion (through automatic focussing or shaking of the camera) might be beneficial in this case, but again these are not ideal motion.

4. Enhancement with General Motion

4.1 Difficulties of General Motion

In the general case, the image formation mechanism is unchanged but other restrictions are relaxed:

- The motion is not uniform over the whole image, or even over a single object, so object shapes can change.

- Object sizes may be changed by changes in camera-to-object distances or focal length changes

- The motion may vary over time; it may be irregular if the camera was held in the hand or mounted on a moving vehicle

- Parts of objects may be hidden ("occluded") by other objects for part of the image sequence

- Motion may continue during the formation of a single image, producing motion blur

- The motion is not planned, and is known only from the image contents

There is no guarantee that image resolution can be improved over that of a single frame. If nothing moved during the sequence, the frames are just repeated samples at the same positions, and pixel-by-pixel processing to reduce noise effects is the only useful way to combine frames.

In most cases, however, the combined pixels from all frames contain samples at many irregularly spaced positions, and might again be used to improve the resolution to a degree that varies from none to that attainable in the ideal case. It will be necessary to accurately register all images to a common coordinate system, and any shortcomings in this step will degrade the final enhanced output.

When the motion of objects has been determined, it may then be possible to determine how the blur in each frame has been affected by that motion, and how much each object has been geometrically distorted by the interaction between motion and sequential scanning.

There is little hope of combining images taken through heat haze, where the distortions are on a small scale and affect few pixels, and so cannot be estimated accurately.

4.2 A Special Single-object Case

A special case of practical interest often arises in analysis of video sequences. Enhancement is required, not of the whole scene, but of a single object. It might be a number plate on a vehicle, a name badge on a person, or something similar. In this case a particular part of the scene must be selected, perhaps by an analyst, but the motion within that part is uniform or simple and there are no occlusions.

In this case the motion may still be irregular and far from the ideal of the previous section. The method for combining frames must still be a general one, but the registration of the frames is easier.

4.3 The Effect of Interlacing

If the sampling of the images was sequential with interlacing, it may no longer be possible to treat a frame as a single, perhaps distorted view of each object. Significant changes in position of an object can occur between sampling events in the two fields. It will then be necessary to treat the fields as separate images, each with only half its lines sampled. Those fields must be registered to each other and to fields of other frames.

Even if nothing moves, odd and even fields will usually differ, because different lines are sampled in each field. For motion to be measured, the lines sampled in one field must be compared with lines not sampled in the next. If a method can be found to register two image regions more accurately than to the nearest pixel, it might also be used to register an even field to an odd field, using just the sampled lines. On the other hand, it might be better to estimate the missing lines first then register two complete images.

The estimation of missing lines in interlaced video is a continuing research area in its own right [de Haan & Bellers, 1998]. Rough estimates can be made just from the known lines in the same field. It may be possible to improve the estimates by bringing in values measured in earlier or later fields, but best results will often require estimating motion locally on the way.

In any case, the accuracy of registration of interlaced fields can depend strongly on the content of the scene. Poor registration can be expected to limit the resolution of the enhanced image.

5. A Method for Enhancing a Slowly Moving Object

5.1 Assumptions

The method assumes that:

- A video sequence is available, which may have been produced by an interlacing system.

- If interlacing was used, each frame has been separated into single-field images.

- A rectangular region of interest (ROI) has been located in the sequence, in which a single object is in slow translational motion (ie without any change in orientation or size). The motion need not be uniform (and the enhancement may be better if it is two-dimensional).

The effects of motion blur will then be neglected.

5.2 Registration

In the following discussion, it is assumed that whole images are to be registered. If only one object within the scene is to be enhanced, the contents of the ROI in each image (frame or field) can be treated as a whole image. The registration will then be valid for the ROI but not necessarily outside it.

The assumption that the motion is purely translational allows the registration to be treated as an image restoration problem [Voss et al, 1999]. With some qualifications for discreteness and pixels near the edges, a translated image is the reference image convolved with an impulse translated from the origin. By an interchange of roles, the translated impulse can be regarded as an unknown image, the translated image as a blurred version of it, and the reference image as the known point spread function (PSF) of the blur. The identification of the translation is then restoration of the impulse followed by location of its peak.

In practice, the translation will not be an integral number of pixels, and the peak will be blurred by the non-integral shift, noise in the images, gain or loss of image features at the edges and any other minor changes between images. Nevertheless, a maximum can be located, and by interpolation it can be found at a non-integral offset.

The restoration and interpolation can be most efficiently done via the Fourier domain, using regularised MMSE restoration and Fourier interpolation.

Let x_{ij}, a_{ij} and y_{ij} be the (i,j) elements of the reference image, the translated impulse and the translated image, all assumed periodic with the same dimensions $m \times n$. Their discrete Fourier transforms will have elements X_{ij}, A_{ij} and Y_{ij}, respectively.

The regularised estimate of A_{ij} is then

$$A_{ij} = \frac{X_{ij}^* Y_{ij}}{|X_{ij}|^2 + \sigma^2 \left\{ \left(\frac{2i}{m}\right)^2 + \left(\frac{2j}{n}\right)^2 \right\}} \quad (1)$$

where "*" denotes complex conjugation, σ controls the degree of regularisation and the expression in braces is based on the use of the Laplacian operator for regularisation. (This is equivalent to a Wiener restoration in which the noise spectrum is white and the image spectrum is a multiple of the reciprocal of the expression in braces. σ^2 is then the signal-to-noise ratio at high frequencies.)

The interpolated impulse is found by extending A with zeros and applying the inverse Fourier transform. (It could alternatively be done by applying the inverse transform first and performing polynomial interpolation.) More precisely, if the $m \times n$ translated impulse image is to be enlarged by a factor $K > 1$, an element of the new transform B_{pq} is given by:

$$B_{pq} = K^2 ST \exp\{i\pi(p/Km - p/m + q/Kn - q/n)\}A_{rs}$$

$$(S, \quad r) = \begin{cases} (1, \quad p), & p < m/2 \\ (1/2, \quad p), & p = m/2 \\ (0, \quad 0), & m/2 < p < Km - m/2 \\ (1/2, \quad p - Km + m), & p = Km - m/2 \\ (1, \quad p - Km + m), & p > Km - m/2 \end{cases}$$

$$(T, \quad s) = \begin{cases} (1, \quad q), & q < n/2 \\ (1/2, \quad q), & q = n/2 \\ (0, \quad 0), & n/2 < q < Kn - n/2 \\ (1/2, \quad q - Kn + n), & q = Kn - n/2 \\ (1, \quad q - Kn + n), & q > Kn - n/2 \end{cases}$$

where pixels are aligned so that the sampled areas coincide, normalisation is assumed to be done by the inverse transform and the special cases are needed to maintain conjugate symmetry of the transform for all even or odd dimensions.

A related approach to the above, phase correlation [Kuglin & Hines, 1975] uses normalised division:

$$A_{ij} = \frac{X_{ij}^* Y_{ij}}{|X_{ij}^* Y_{ij}|} \quad (2)$$

Both these methods were tested on clean and noisy image sequences. The second was found to be far more sensitive to noise and poorer at estimating non-integral displacements, so it was not considered further. The first functioned well for an enlargement of the impulse by a factor around 10 and a fixed large value for σ (say 10^5), with the following problems:

1. If some parts of the scene moved relative to the rest between images, the displacement detected was a mean of local displacements rather than a mode, with higher weight given to regions of stronger contrast. This was important when the images were assumed periodic, and opposite edges "wrapped around" but did not match. The line of mismatch behaved like a stationary object and the displacement was then underestimated. (For example, if ground and sky appeared, they met at the horizon and also where the top met the bottom. The horizon could move, but the top and bottom never did.)

2. For an enlargement by an even value K, zero shifts could not be detected as zero. This happened because the pixel representing zero shift in the translated pulse image was scaled into a K by K block, at the centre of which one of four pixels representing shifts of $\pm 1/2k$ pixels was chosen. (If $K = 10$ and an error of ± 0.05 pixels is adequate, this problem is mainly aesthetic).

Problem 1 was suppressed by subtracting a 5x5 local mean from each pixel value, (in effect applying a high-pass filter) and regarding the image as extended outside its boundaries by reflection. (In that way, most pixels near the edges became zero and matched their counterparts on the opposite edge.) This step had an unwanted side-effect in the case where missing lines of a single interlacing field had been estimated by interpolation, especially when the interlacing was in four fields, because the changes of gradient ("creases") across the lines with real samples could be falsely recognised as stationary objects. A further step, vertical low-pass filtering with a kernel

$$(1/16 \ 2/16 \ 3/16 \ 4/16 \ 3/16 \ 2/16 \ 1/16)$$

was added before mean subtraction to smooth out the creases. (Applying these filters in the Fourier domain would probably be slower.)

Problem 2 was avoided by using 11 rather than 10 as the enlargement factor, but could also be removed by treating the translated pulse image as having its origin at the centre of the zero-shift pixel. This eliminates the complex exponential term from the Fourier interpolation formula and allows an even factor such as 10.

With these problems taken care of, and with fixed $\sigma = 10^5$, the impulse restoration method above was found very effective for estimating a single displacement between two video images, to within 0.1 of the pixel size, even when applied to fields reconstructed from every second (or fourth) line.

An alternative method for registering fields of interlaced sequences was to use only the known lines of each field without interpolation, treating them as full images. The vertical displacement estimates are then doubled (or quadrupled) and offset to take account of the different starting positions of different fields. This alternative was tested on one sequence with two-field interlacing. Compared to the method described above, it had the disadvantage that displacements were estimated to the nearest 2/11 of a pixel instead of 1/11. Otherwise it produced similar results. (A better approach might be to enlarge vertically by 21 at the Fourier interpolation step, or use local polynomial interpolation.)

5.3 Enhancement

Under the assumptions of Section 5.1, within the ROI, low-resolution digitisation of a continuous scene has been performed for different displacements of the object of interest, and possibly under noisy conditions. An estimate of one higher-resolution digitisation of the same scene, with a single representative displacement, is required.

If the result is to show detail at the size of its pixels, there should be enough information to determine an independent value for each pixel. This requires that there be at least as many input pixels as output pixels. For a given number of input frames of the same scene, the resolution improvement is strictly limited by this rule. (For the ideal motion case above it was satisfied by the design of the frame sequence.)

In practice it will not be possible to estimate a continuous image and derive discrete samples from it. Some further simplifying assumptions are now made:

- The higher-resolution sample spacing of the output image is an integer sub-multiple $1/M$ of the lower-resolution spacing of the input images. (The case $M = 1$ is useful; it allows noise reduction by averaging over differently aligned images when resolution enhancement is not needed.)

- The output image is registered with a specified input image, covering the same area with M^2 pixels in the output image for every pixel in the input image.

- The sample area of a higher-resolution output pixel covers the whole pixel area. (This is a valid design choice for a fictional high-resolution camera.)

- The sample area of the lower-resolution input pixel covers the whole pixel area. (This may be false, and other smaller sizes need to be considered.)

- The underlying continuous image has no details other than what will be shown in the discrete output image; rather, it is constant over each sample area.

Under these assumptions, each input sample, once the displacements between images are known, can be related to an area of the continuous image that overlaps higher-

resolution pixels by known amounts. (See figure 1.) Its value is a weighted mean, with known weights, of the unknown values of the higher-resolution pixels, corrupted by noise. This approach has been taken by various authors (eg [Elad & Feuer, 1999a]).

Higher-resolution samples (M=3)

Lower-resolution sample
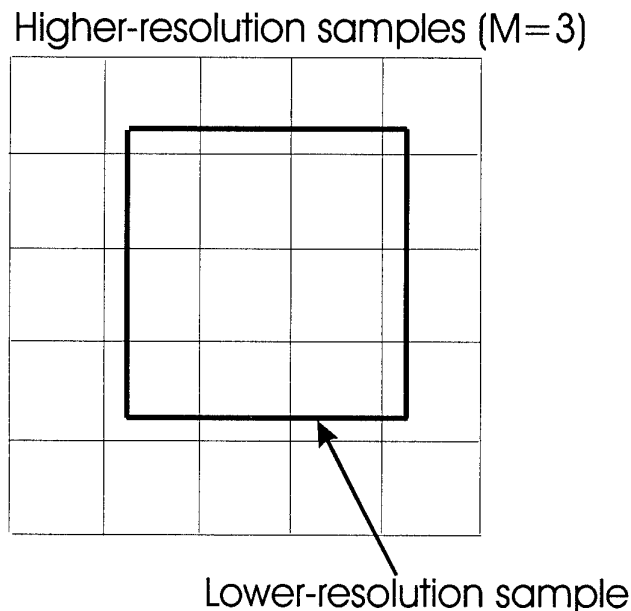
Figure 1. Relation of a lower-resolution sample to higher resolution samples

Let the unknown higher-resolution pixel values be arranged in some convenient order as a vector $\mathbf{x}$, and the lower-resolution pixel values from the input images as a vector $\mathbf{y}$. Then the lower-resolution sampling process is described by

$$\mathbf{Ax} + \mathbf{n} = \mathbf{y} \quad (3)$$

where $\mathbf{A}$ is a (very sparse) matrix of the weights in the lower-resolution sampling process and $\mathbf{n}$ is a noise vector. A regularised solution of (3) for $\mathbf{x}$ minimises

$$E = |\mathbf{Ax} - \mathbf{y}|^2 + \lambda|\mathbf{Lx}|^2 \quad (4)$$

where $\mathbf{L}$ applies a high-pass operation (typically the 2-D Laplacian at non-boundary pixels) to $\mathbf{x}$ and λ is a constant that controls the amount of regularisation. The solution is found by solving

$$\left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{L}^T\mathbf{L}\right)\mathbf{x} = \mathbf{A}^T\mathbf{y} \quad (5)$$

which will be written

$$\mathbf{Cx} = \mathbf{b}. \qquad (6)$$

Once $\mathbf{b}$ is computed, solution for $\mathbf{x}$ requires a numerical method. A well-tried choice is a pre-conditioned conjugate gradient method [Golub & Van Loan, 1996, Section 10.2]. Expressed in pseudo-code, the method is:

```
x = initial guess
ε = maximum relative error
P = pre-conditioning matrix
r = b - Cx
β = 0
p = 0
DO until |r| < ε|y|
      z = P⁻¹r
      q = rᵀz
      IF not first time
            β = q/q'
      ENDIF
      p = z + βp
      q' = q
      α = q/pᵀCp
      r = r - αCp
      x = x + αP
ENDDO
```

Here $\mathbf{P}$ is chosen to approximate $\mathbf{C}$ but to allow easy multiplication by $\mathbf{P}^{-1}$. The diagonal part of $\mathbf{C}$ is a suitable choice.

$\mathbf{C}$ is only used to construct $\mathbf{P}$ and to pre-multiply vectors. In these roles it can be applied indirectly using only lists of which lower-resolution samples depend on which higher-resolution samples (equivalent to $\mathbf{A}$), and the definition of the Laplacian operator. It need not be constructed explicitly at any stage. Likewise, $\mathbf{b}$ can be constructed without explicitly computing $\mathbf{A}$. $\mathbf{P}$ is diagonal, so its diagonal elements can be computed once and stored. Thus full advantage is taken of the sparsity of $\mathbf{A}$.

The error tolerance ε is typically set to 10^{-5}. The value of λ is assumed to be chosen by trial and error: large values cause loss of image detail, while small values give noisy results, so λ should be set to the smallest value for which noise in the output is not distracting.

If different resolution gains M are to be compared, a fixed value of λ may not be appropriate. Suppose that M is made large enough for all the detail available from the input to appear in the output. If M is further increased, the effect on the output should be one of interpolation, without any change in the amplitude of detail. In equation (3),

$\mathbf{y}$ will not change in dimension or value, so $\mathbf{n}$ and $\mathbf{Ax}$ (as estimated) should do approximately the same. The vector $\mathbf{x}$ has similar values but the number of them has increased as M^2, so the number of columns in $\mathbf{A}$ increases as M^2 and its element values should decrease as M^{-2}.

Now consider the balance of equation (5). On the right-hand side, the changes to $\mathbf{A}$ in $\mathbf{A}^T\mathbf{y}$ produce more rows (varying as M^2) and smaller values (varying as M^{-2}). On the left, the term $\mathbf{A}^T\mathbf{Ax}$ changes the same way because $\mathbf{Ax}$ is unchanged. Apart from some minor asymmetries at the boundaries, $\mathbf{L}$ is square and symmetric and finds second differences. The features being processed are now scaled up as M, so the second differences are reduced in value as M^{-2}. Multiplying by $\mathbf{L}^T$ does more or less the same again, so the term $\lambda\mathbf{L}^T\mathbf{Lx}$ has more rows (as M^2) with values reduced as M^{-4}. To preserve the balance as M changes, a further factor of M^2 should be introduced in this term, and this can be done by keeping λ fixed and using $M^2\lambda$ in the restoration.

Test results with varying M confirm this prediction, and suggest that 1 is a good first guess for λ, followed by 10 or 0.1 .

5.4 Computer implementation

The method of Section 5.3 has been coded in ANSI C, in two programs called *imgetshift* and *imergesxnjcg*. These operate on images in the PGM (8-bit grey) format and produce a final output in a similar floating-point format for separate conversion to PGM.

5.4.1 *imgetshift*

In the first step, *imgetshift* reads a list of input images, in which the first is regarded as the reference image. It reads and filters the reference image and finds its Fourier transform. It then reads each of the other images, filters it, finds the Fourier transform, applies equation 1 with fixed σ, and applies the inverse transform. To allow interpolation, it extends with zeros vertically before applying the vertical transform, but then extends and applies the transform horizontally only one line at a time – only the location of the maximum is needed. The final output of this program is the list of input images, each marked with its displacement from the reference image, estimated to the nearest 1/11 pixel.

At the head of the list must appear the location and size of the ROI and the number of images. Only the portion of each image within the ROI is processed. The same ROI and count appear in the output list.

Additional information about each image, namely whether it is a single interlacing field and whether odd, even etc., is passed through unchanged and does not affect the processing in this step. Separation of fields is assumed done in an earlier step.

5.4.2 *imergesxnjcg*

In the second step, *imergesxnjcg* reads the list produced by *imgetshift*. The user also specifies the regularisation constant λ, the tolerance for the conjugate gradient method, a limit on the number of iterations and a zoom factor M. The ROI in the first image (the reference image), enlarged by M, specifies the size and range of the output image.

The program then reads each image and determines from the displacement of that image which pixels fall into the ROI in the reference image. Each such pixel (if it is in a valid line for any interlacing specified) is then recorded as a sample of relevant output pixels in the ROI, and a list equivalent to the sparse matrix **A** is built up. The pre-conditioned conjugate gradient method described in Section 5.3 is used to find the output.

Variants of this program exist for the alternative initial guesses of zero and a simple zoomed version of the ROI of the reference image. For $M = 1$ the more complicated guess saves only an iteration or two, but for $M > 1$ and interlaced images time savings up to 40% have been obtained.

Several tens of iterations are usually required for a tolerance of 10^{-5}. The most slowly decaying errors appear to be associated with the boundary of the output image and uneven sampling. Uneven sampling can be caused in turn by special cases in the motion of the camera, by interlacing and by the absence of samples from near part of the boundary of the output in images other than the reference image. (Convergence can then be faster if the ROI does not come too close to the boundary.)

5.5 A variant for images with interference

In some tests of this method, poor results were partly attributable to variations in image brightness that might be the result of electrical interference to the video signal. This subsection considers special treatment for the problem, regardless of its actual cause.

5.5.1 Nature of the problem

Consider a sinusoidal signal of possibly varying frequency added to the brightness during the scanning process. If it has a frequency near an integer multiple of one cycle per line scan, vertical or diagonal fringes will result. If its frequency is one cycle for every few line scans, horizontal fringes will result. For frequencies of the order of one cycle for every frame, the brightness variation may not be noticeable within a frame but will be distinct from frame to frame. (Such effects may be seen in the picture of a

broadcast television receiver when reception is poor. Frame-to-frame changes could also be the result of continuous automatic exposure adjustment by the camera itself.)

In all cases, the brightness of the same small object in the scene can vary from frame to frame. If interlacing is used, the same variation may occur between fields in one frame and will be more noticeable within a frame

When resolution is to be enhanced by combining several frames (or fields), pixel values from different frames are interpreted as samples of a high-resolution image at slightly different positions. This may lead the enhancement method to interpret the interference effects as fine detail in the enhanced image, and to amplify them.

(Note that under the assumptions of Section 5.3 with no interlacing, the problem is suppressed. The variation from frame to frame cannot be accounted for by fine detail in the enhanced image, for the required detail would have a period of 1 input pixel, or M output pixels, and would have the same mean over every $M \times M$ sample. The variation can only be interpreted as noise in the input. If the sample size and spacing are assumed different, or the input images are interlaced, the output is affected.)

Because the spatial wavelengths involved are often of the same order as object sizes, there is little hope of filtering out the interference once it has corrupted the image. Rather, the resolution enhancement process must be modified to remove fine details implied only by slowly varying grey level differences between input images.

5.5.2 Possible solution

Suppose complementary low-pass and high-pass filters are applied to different copies of each image, and designed with a cutoff frequency low enough for features a few pixels in size to be preserved in the high-pass outputs, but high enough so that interference fringes and broader changes appear only in the low-pass outputs.

The low-pass outputs are now combined using the method of Section 5.3 with a larger value for λ so that fine details implied by grey-level differences will be suppressed. The result is like the mean of the images after registration. In practice, similar results could be achieved by using the raw input images and the larger λ without applying the low-pass filter.

The high-pass outputs are combined using the same method with a smaller λ value so that fine details are preserved and differences between images due to the slow motion or interlacing effects are correctly interpreted as fine details, free of interference problems.

The two enhanced outputs are added to give a result that responds to fine detail in the differences between images, but not the coarser features of those differences. Further analysis is required of the frequency response of this method for different high-pass

filter and λ combinations, but tests on synthetic images of sinusoids and real images with interference suggest that using a simple 9-point high-pass filter, $\lambda = 0.3$ for the raw input images and $\lambda < 0.1$ for their high-pass components gives satisfactory frequency response and greatly reduced artefacts from the interference.

The use of a reference frame for brightness and all frames for edge information has been considered by Chiang & Boult [1997b]. Another approach suited to changes in illumination between frames was taken by Granrath & Lersch [1998].

5.6 Testing

Shift estimation and enhancement were tested using images with synthetically reduced resolution and real interlaced video sequences. (The examples presented here do not need the treatment for interference discussed in the previous sub-section.)

5.6.1 Synthetic sequence

Figure 2(a) shows the "cameraman" test image, with grey level range 0-255, trimmed to 240x240 pixels. Reduced-resolution images were generated by averaging grey levels over 3x3 pixel blocks, starting at nine different offsets (0, 1 or 2 in x, and 0, 1 or 2 in y) from the top-left corner. Figure 2(b) shows one of these images enlarged to the original size; as an estimate of the original it had an RMS error of 12.73.

2(a) Original   2(b)Reduced-resolution copy

2(c) Resolution gain 1

2(d) Resolution gain 2

2(e) Resolution gain 3

Figure 2. Enhancement from reduced-resolution copies of a single image

The relative offsets of the nine images were then estimated and enhanced single images were prepared with resolution gains of 1, 2 and 3. Figures 2(c)-2(e) show these enhanced images made the same size as the original by spline interpolation so that resolution improvement can be judged by eye and measured. The respective RMS errors were 10.69, 7.25 and 6.41. A gain of 3 gave better results than a gain of 2, but not all the details of the original could be recovered.

5.6.2 Real interlaced video sequence

Figures 3(a)-3(b) show frames 1 and 3 of the longer "flower garden" 720x485 interlaced video sequence. In this sequence the camera moves to the right and away from the scene, producing parallax and slight scale changes not equivalent to a simple shift.

3(a) Frame 1 with ROIs

3(b) Frame 3

3(c) Enhancement of window

3(d) Enhancement of flowers

3(e) Enhancement of tree
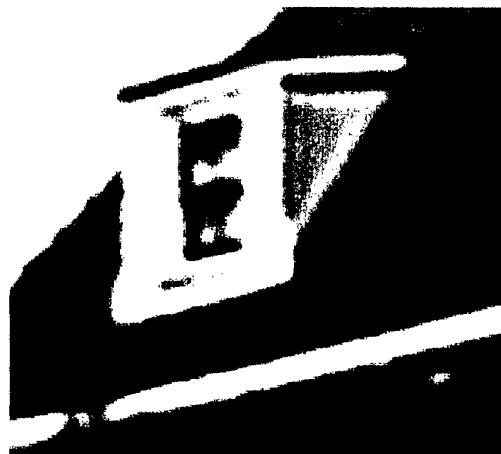
3(f)Interpolation vs. enhancement of window

Figure 3.Enhancement from an interlaced video sequence for selected Regions Of Interest

Three small ROIs were chosen: near the attic window, within the flower bed, and on the tree trunk, as shown in Figure 3(a). Shifts were estimated for each ROI for the six fields of the first three frames relative to the first (even) field of frame 1. Three

enhancements were then performed, each using the shift estimates for a different ROI, but applied with a resolution gain of 3 to a 200x200 area containing all three ROIs. Figures 3(c)-3(e) show the respective results. Figure 3(f) shows detail of the window in frame 1 after smooth (cubic) interpolation and after enhancement. No "true" image is available for comparison, but each result shows a visual improvement within its relevant ROI, and this improvement is beyond what zooming can achieve.

6. Methods for more general motion

6.1 Registration

Registration of images requires a much more sophisticated approach when objects are moving without restriction, and there may be occlusions, rotations and size changes. It will be necessary to recognise different scales of detail, so that if for example an object with spots is being tracked, different spots are not confused. If one image is chosen as a reference image, and objects are to have the same positions in the enhanced output that they have in that image, there will usually be features in that image that are not visible in other images and cannot be enhanced. At the same time there will be features visible only in other images that cannot contribute to the output. It may still be necessary to assume that movements are small between consecutive images, just to identify the movements.

Much work has been done on using different approaches for registration. For a survey see [Brown, 1992].

In the present work, one method, based on the use of complex wavelets, was fully considered. This method was developed by [Magarey & Kingsbury, 1996], and coded as a set of macros for Matlab, under the name *cdwtgui*.

cdwtgui accepts a pair of images and attempts to relate each pixel in the second to a pixel in the first, with sub-pixel precision, as a displacement vector. To set up an initial guess, the user selects two points visible in both images and identifies them to the program by pointing to them in each image. The program also requests the value of a parameter that controls the degree of smoothness of the displacement vector, and thus the relative importance of smoothness and accuracy of match.

By the use of wavelet analysis, the program compares the images for details at different scales. The best match at each scale, starting with the coarsest, is used as the initial guess for the next finer scale. The finest-scale result, further interpolated if necessary, gives the X- and Y-displacements from each pixel in the second image to the matching pixel in the first.

The method is quite fast, and its accuracy well inside objects is good, but the use of a single smoothness parameter means that the smoothing that occurs within a single

object must also affect the edges separating that object from other objects that may be moving differently. There is then a trade-off between accurately estimating the movement of an object at its centre and approximately estimating that movement near its edges. Further difficulties are the requirement that image dimensions be divisible by large powers of two and the lack of any means to extrapolate a match of part of one object in both images into other areas to which the same object might extend.

A fully satisfactory image-matching procedure needs to detect the boundaries of differently moving objects, so that the displacement estimated for each pixel uses matches for nearby pixels within the same object but not those belonging to different objects. Any smoothness condition must therefore allow occasional discontinuities across object boundaries.

The methods of Section 5, using a ROI within a single object, could be expected to match that object up to its boundaries, even outside the ROI, if it is moving uniformly without scale change, rotation or distortion against an unrelated background. (This method can be tested using the implementation of Section 5.4 if the ROI is altered between the two stages by simple editing of an intermediate file. Examples appear in Figure 3.)

A method due to Thévenaz [1998] was briefly tested using publicly available software [Thévenaz, 1997]. This allows for uniform translation, rotation and shearing between two images, with some limitations on the extent of changes, and is coded as a set of C functions. Its output includes the second image transformed to match the first and a list of transformation coefficients, which could be used to relate pixel coordinates in the two images with sub-pixel precision. The method apparently worked well, and could be used to extend the method of Section 5.4 so that it can be applied to a single object that deforms over time.

A method for estimating affine transformations between frames was included in the work of Granrath & Lersch [1998].

6.2 Enhancement

In the case of affine changes only between frames, the method of Granrath & Lersch [1998] is well worth considering, if only for speed of operation. In the case of more general motion the method is not applicable, but some other form of interpolation combined with their use of the POCS approach may be applicable.

An image matching program such as *cdwtgui* can be used to produce an enhanced single image from a video sequence, with motion still slow but otherwise general, as follows:

• A reference image is chosen to control what will be visible in the enhanced output, and a zoom factor M is selected.

- The image matching program is used to compare each other image with the reference image, to define the location in the reference image of every pixel of every other image. Then every pixel in every image has a known location in the reference image.

- Each input pixel value (if in a valid line for a single-field image, or any line for a full-frame image) is considered a sample over a continuous image. The continuous image is to be reconstructed and re-sampled to give a higher-resolution output image, and is assumed to be constant over each output sample. Each input pixel value is thus expressed as a combination of output pixel values. (Here "input" and "output" are relative to the restoration process, not to the camera.)

- The method of Section 5.3 can now be applied.

Strictly, the assumptions made in Section 5.3 require that the overlap of each input pixel with the (supposed constant) output pixel areas must be computed. The approach has only been tested under the approximation that each input pixel has the location found by the registration procedure, but a size $M \times M$ output pixels and the same orientation as the output pixels. Whether the approximation is made or not, the uneven sample spacing may lead to artefacts in the presence of the interference described in Section 5.5.1, even when there is no interlacing.

If the motion is not slow, the set of output pixels considered to contribute to an input pixel will be increased by the motion of the image in the image plane during the time the sample was being formed. This motion must in general be estimated from the positions in the output of the same pixel in consecutive frames (or fields), and the pixel combination is effectively a convolution of the sample area with a curve representing the motion. Examples were seen during this study of hand-held video camera sequences in which the motion changed noticeably between consecutive fields of the sequence, so proper compensation for all motion will be at best difficult.

The case of vibrational motion has been considered by [Stern *et al*, 1999]. There it is shown that careful selection of a subset of frames is needed for a good result.

If the motion is not slow but is uniform, the samples will approximate those assumed for slow motion convolved with a straight line segment of uniform density. Then the effect of motion could be removed by first performing the enhancement assuming no motion, then performing restoration for motion blur separately.

6.3 Computer implementation

The final stage of enhancement in Section 6.2 has been implemented in the C language as a program *imergexnjcg*, and tested in combination with *cdwtgui* and with separate

conversion steps for the input image pairs passed to *cdwtgui* and for the displacement fields passed back.

In the final step, *imergexnjcg* reads a list of input images, which specifies for each which interlacing field it contains (if any) and in which image the corresponding displacement fields are kept. The displacement image can be left unspecified for an input image, indicating that the displacements are all zero; this is done for the reference image itself, but the reference image need not appear in the list. The user also specifies the ROI, the regularisation constant λ, the tolerance for the conjugate gradient method, a limit on the number of iterations and a zoom factor M. The ROI in the reference image, enlarged by M, specifies the size and range of the output image.

The program then reads each input image and determines from the displacements of that image which pixels are in the correct lines for any interlacing specified and map into the ROI in the reference image. Each such pixel is then recorded as a sample of relevant output pixels in the ROI, and a list equivalent to the sparse matrix **A** of Section 5.3 is built up. The pre-conditioned conjugate gradient method described in that section is used to find the output.

6.4 Testing

A number of alternatives to the conjugate gradient method were tested, mostly on irregularly placed samples from one-dimensional synthetic data. These methods were based on approaches that had apparently performed well in other applications, but they did not produce fast and reliable results as expected. They are summarised in Appendix A, mainly to document their failure.

6.4.1 Enhancement from random samples

Enhancement from randomly-placed samples was tested by generating 2x2 or 3x3 pixel samples of synthetic test images at random (and non-integral) locations and treating them as pixel values from low-resolution images with known displacements back into the test image. The distribution of the samples was not uniform, but designed so that the density of samples (per test image pixel) varied from much less than one to greater than one.

Figure 4(a) shows one of the test images, 64x64 pixels. Random sample positions were generated with a probability density proportional to the distance from the top edge until their density at the bottom was 2 samples per pixel. 2x2 samples were taken at these positions. Figures 4(b) and 4(c) show the reconstructions for regularisation constant values 0.0001 and 0.01 respectively.
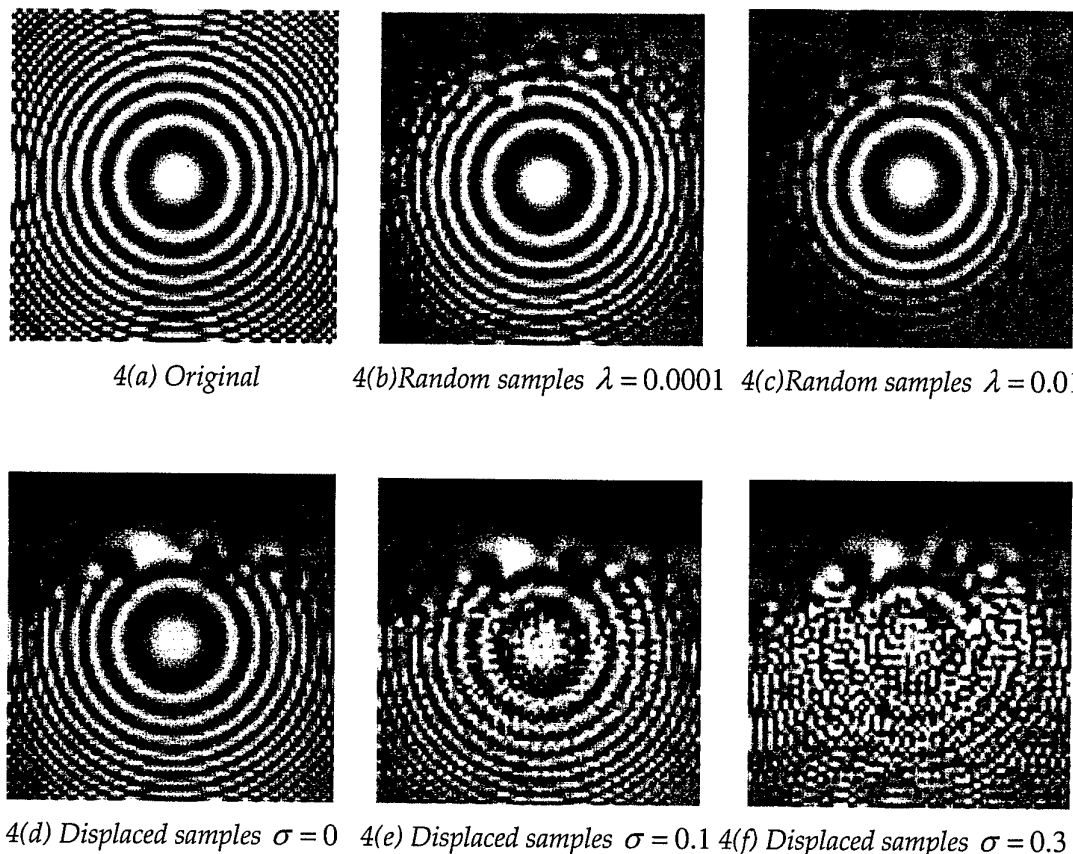
4(a) Original　　　4(b)Random samples $\lambda = 0.0001$　4(c)Random samples $\lambda = 0.01$

4(d) Displaced samples $\sigma = 0$　4(e) Displaced samples $\sigma = 0.1$ 4(f) Displaced samples $\sigma = 0.3$

Figure 4. Image reconstruction from random 2x2 samples.

The test was varied to add a gaussian random error in x and y to each sample location between sample generation and reconstruction. In this case the probability density increased as the cube of the distance from the top of the image, and the sample density was 8 samples per pixel at the bottom, 1 sample per pixel half way down. Figures 4(d) to 4(f) show the results for RMS errors of 0, 0.1 and 0.3 pixels in each coordinate. In each case the regularisation constant was 0.00001.

These results show that sample densities as low as two per output pixel are enough to give a good reconstruction, but that the presence of registration errors of a few tenths of a pixel may require more samples, from more input images.

6.4.2 Enhancement from irregular motion

The implementation of Section 6.3 was applied to real video sequences, with and without interlacing. The *cdwtgui* program did not give accurate displacements right up to the edges of differently moving objects, but tended to leave smooth transitions at those edges. The resulting poor displacement values were reflected in poor output image quality there.

Figure 5. Enhancement of a whole image with differently moving objects

Figure 5 shows part of the "garden" scene (compare Figure 3), enhanced with resolution gain 2 from two frames (four fields) . Far more of the image was satisfactorily enhanced than by using a fixed shift (as in Figures 3(c) to 3(e)), but if one object was of interest, less of it would be restored than by the fixed-shift method.

7. Conclusions

Methods have been developed for the extraction of higher-resolution images from short video sequences. These include:

- Simple re-ordering of data and de-blurring, when camera positions are suitably chosen in advance
- Shift estimation by a Wiener filtering approach, followed by regularised minimum mean square error reconstruction, when only a single object is of interest and it moves slowly without change of size or orientation
- Registering images by a wavelet-based approach, followed by regularised reconstruction, when several objects are to be extracted or motion is more general.

Many alternatives were considered in the choice and implementation of these methods, as described in Appendix A, but gave slow or unsatisfactory results. Some of the other methods referred to in Section 2 could be further considered.

The chosen methods can allow for interlaced video scan and some kinds of interference patterns in the images. The effects of varying the amount of data available and poorly estimating object displacements have been considered.

The improvement in resolution that can be obtained is limited by many things, some of which affect image restoration work in general:

1. The amount of input data. For a resolution gain of 3, for example, 3x3=9 frames are needed, even if conditions are ideal. Some test results suggested that twice this many frames might be needed.
2. Noise in the input. This will be amplified by attempts to improve resolution, and can only be overcome by using more regularisation (with loss of resolution) or more input data.
3. The size of the area sampled for each pixel in the camera, relative to the pixel spacing. This limits the gain to 2 for accurate restoration (in the presence of weak noise) when the focal plane is fully covered by pixels. Smaller samples (more likely in real cameras) allow a higher gain if there are enough input frames. Changes of subject distance or zoom also help to overcome this problem.
4. Blur caused by an out-of-focus or moving camera, even if the extraction method takes full account of it. The effect is analogous to that of pixel sample size, and might be reduced if the focus or motion varies.
5. Fortuitous alignment of frames. The camera or subject must move enough so that pixels are well scattered over the subject. Pure horizontal or vertical motion may only improve resolution in the direction of motion, while slight rotation or change of subject distance on its own may leave patches of low resolution. In some cases this may be overcome by using more frames.
6. Poor registration. Errors larger than 0.1 pixels in the positioning of input pixels relative to the (possibly smaller) output pixels were shown to lead to poor results. This imposes a minimum size on output pixels.
7. Interlacing in a video camera. The absence of half the lines in each field may introduce errors into the registration process.

The main limitation in tests appeared to be the quality of registration, particularly near the boundaries of differently moving objects. Further study of new and existing registration methods could improve the handling of general object motion, leading to better resolution enhancement of whole objects.

References

Alam M.S., Bognar J.G., Hardie R.C. & Yasuda B.J. (1997) High resolution infrared image reconstruction using multiple, randomly shifted, low resolution, aliased frames, *Proceedings SPIE Conference on Infrared Imaging Systems: Design, Analysis, Modeling and Testing VIII*, SPIE **3063**, 102-112.

Albertz J. & Zelianeos K. (1990) Enhancement of satellite image data by cumulation, *ISPRS Journal of Photogrammetry and Remote Sensing*, **45**, 161-174.

Armstrong E.E., Hayat M.M., Hardie R.C., Torres S.N. & Yasuda B. (1999) Non-uniformity correction for improved registration and high-resolution image reconstruction in IR imagery, *Proceedings SPIE Conference on Applications of Digital Image Processing XXII*, SPIE **3808**, 150-161.

Avrin V. & Dinstein I. (1997) Restoration and resolution enhancement of video sequences, *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, **IV**, 2549-2552.

Bascle B., Blake A. & Zisserman A. (1996) Motion deblurring and super-resolution from an image sequence, *Proceedings European Conference on Computer Vision*, **2**, 573-582.

Brown L.G. (1992) A Survey of Image Registration Techniques, *ACM Computing Surveys*, **24**(4), 325-376.

Cheeseman P., Kanefsky B., Kraft R., Stutz J. & Hanson R. (1994) Super-resolved surface reconstruction from multiple images. Technical Report FIA-94-12, NASA Ames Research Center.

Chen D. & Schultz R.R. (1998) Extraction of high-resolution video stills from MPEG image sequences, *Proceedings IEEE International Conference on Image Processing*, 465-469.

Chiang M. & Boult T.E. (1997a) Local blur estimation and Super-Resolution, *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 821-826.

Chiang M. & Boult T.E. (1997b) Imaging-Consistent Super-Resolution, *Proceedings ARPA/ISO Image Understanding Workshop*, **II**, 1439-1445, May 1997.

Daubechies I. (1992) *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics., **61**, Society for Industrial and Applied Mathematics., Philadelphia

de Haan G. & Bellers E.B. (1998) Deinterlacing – An Overview, *Proceedings IEEE*, **80**(9), 1837-1857.

Elad M. & Feuer A. (1977) Restoration of a single superresolution image from several blurred, noisy and undersampled measured images, *IEEE Transactions on Image Processing*, **6**(12), 1646-1658.

Elad M. & Feuer A. (1999a) Superresolution Restoration of an Image Sequence: Adaptive Filtering Approach, *IEEE Transactions on Image Processing*, **8**(3), 387-395.

Elad M. & Feuer A. (1999b) Super-resolution Reconstruction of Image Sequences, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(9), 817-834.

Gillette J.C., Stadtmiller T.M. & Hardie, R.L. (1995) Aliasing reduction in staring infrared imagers utilizing subpixel techniques, *Optical Engineering*, **34**(11), 3130-3137.

Golub H.G. & Van Loan C.F. (1996) *Matrix Computations*, Ed. 3, John Hopkins University Press.

Granrath D. & Lersch J. (1998) fusion of images on affine sampling grids, *Journal of the Optical Society of America*, **15**(4), 791-801.

Hardie R.C., Cain S., Barnard K.J., Bognar J., Armstrong E. & Watson E.A. (1997a) High resolution image reconstruction from a sequence of rotated and translated infrared images, *Proceedings SPIE Conference on Infrared Imaging Systems: Design, Analysis, Modeling and Testing VIII*, SPIE **3063**, 113-124.

Hardie R.C., Barnard K.J. & Armstrong E. (1997b) Joint MAP registration and high-resolution image estimation using a sequence of undersampled images, *IEEE Transactions on Image Processing*, **6**(12), 1621-1633.

Hardie R.C., Tuinstra T.R., Bognar J., Barnard K.J. & Armstrong E. (1997c) High resolution image reconstruction from digital video with global and non-global scene motion, *Proceedings IEEE International Conference on Image Processing*, **I**, 153-156.

Hildebrandt J.W. & Newsam G.N. (1990) Subpixel Resolution through Dithering, *Proceedings 2nd IASTED International Symposium on Signal Processing and its Applications*, 515-518, Gold Coast 1990.

Irani M. & Peleg S. (1991) Improving Resolution by Image Registration, *CVGIP: Graphical Models and Image Processing*, **53**(3), 231-239

Irani M. & Peleg S. (1993) Motion analysis of image enhancement, resolution, occlusion and transparency, *Journal of Image Communication and Representation*, **4**(4), 324-355.

Kuglin C.D. & Hines D.C. (1975) The Phase Correlation Image Alignment Method, *Proceedings IEEE Conference on Cybernetics and Society*, 163-165.

Lai S. & Cui Y. (1999) Total variational blind image restoration from image sequences, *Proceedings SPIE Conference on Visual Communications and Image Processing '99*, SPIE **3653**(1), 96-105.

Lai S. & Vemuri B.C. (1997) Physically Based Adaptive Preconditioning for Early Vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(6), 594-607

Leung W.V. & Lane R.G. (2000) Joint blind deconvolution and registration of a sequence of undersampled images, *Proceedings of Image and Vision Computing New Zealand 2000*, 50-55.

Lorette A., Shekarforoush H. & Zerubia J. (1997) A super-resolution with adaptive regularization, *Proceedings IEEE International Conference on Image Processing*, I, 169-173.

Luca P., Piergorgio N. & Oliviero B. (1997) A resolution improvement technique for two-dimensional detectors, *Proceedings Europto Conference on New Image Processing Techniques: Algorithms, Methods and Components II*, SPIE 3101, 83-94, 1997.

Magarey J. & Kingsbury N. (1996) Motion Estimation Using Complex Wavelets, *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, 2371-2374, May 1996.

Patti A.J., Sezan M.I. & Tekalp A.M. (1994) High-resolution Image Reconstruction from Low-resolution Image Sequence in the Presence of Time-varying Motion Blur, *Proceedings IEEE International Conference on Image Processing*, I, 343-347.

Patti A.J., Sezan M.I. & Tekalp A.M. (1997) Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time, *IEEE Transactions on Image Processing*, 6(8), 1064-1076.

Patti A.J. & Altunbasak Y. (1998) Artifact reduction for POCS-based super resolution with edge adaptive regularization and higher-order interpolants, *Proceedings IEEE International Conference on Image Processing*, 217-221.

Patti A.J. & Altunbasak Y. (2001) Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher order interpolants, *IEEE Transactions on Image Processing*, 10(1), 179-186

Richardson W.H. (1972) Bayesian-based iterative method of image restoration, *Journal of the Optical Society of America*, 62, 55-59

Sale D., Schultz R.R. & Szczerba R.J. (2000) Super-resolution enhancement of night vision image sequences, *Proceedings IEEE International Conference on Systems, Man & Cybernetics*, 3, 1633-1638.

Schultz R.R. & Stevenson R.L. (1996) Extraction of high-resolution frames from video sequences, *IEEE Transactions on Image Processing*, 5(6), 996-1011.

Schultz R.R. & Stevenson R.L. (1998) Estimation of subpixel-resolution motion fields from segmented image sequences, *Proceedings SPIE Conference on Sensor Fusion: Architectures, Algorithms, Applications II*, SPIE 3376, 90-101.

Stark H. & Oskoui P. (1998) High-resolution image recovery from image-plane arrays, using convex projections, *Journal of the Optical Society of America (A)*, **6**(11), 1715-1726.

Stern A., Kempner E., Shukrun A. & Kopeika N.S. (1999) Restoration and resolution enhancement of a single image from a vibration distorted image sequence, *Proceedings SPIE Conference on Applications of Digital Image Processing*, SPIE **3808**, 224-233

Sweldens, W. (1995) The Lifting Scheme: A New Philosophy in Biorthogonal Wavelet Constructions, *Proceedings SPIE Conference on Wavelet Applications in Signal and Image Processing*, San Diego, SPIE **2569**(1), 68-79

Tekalp A.M., Ozkan M.K. & Sezan M.I. (1992) High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration, *Proceedings IEEE International Conference on Signal Processing and its Applications*, **III**, 169-172.

Thévenaz P. (1997) Computer software regFlt3d, Biomedical Imaging Group, Swiss Federal Institute of Technology Lausanne,
URL - http://bigwww.epfl.ch/thevenaz/registration/index.html

Thévenaz P., Ruttimann U.E. & Unser M. (1998) A Pyramid Approach to Subpixel Registration Based on Intensity, *IEEE Transactions on Image Processing*, **7**(1), 27-41.

Tull D.L. & Katsaggelos A.K. (1995) Iterative restoration of fast moving objects in dynamic image sequences, *Proceedings SPIE Conference on Visual Communications and Image Processing*, SPIE **2501**(2), 1088-1098.

Ur H. & Gross D. (1992) Improved resolution from subpixel shifted pictures, *CVGIP: Graphical Models and Image Processing*, **54**(2), 181-186.

Voss K., Ortmann W. & Baumbach T. (1999) Shift detection by restoration, *Pattern Recognition*, **32**, 2067-2068.

Yang K. & Murty K.G. (1992) New Iterative Methods for Linear Inequalities, *Journal of Optimization Theory and Applications*, **72**(1), 163-185

Youla D.C. & Webb H. (1982) Image Restoration by the Method of Convex Projections: Part 1 – Theory, *IEEE Transactions on Medical Imaging*, 1(2), 81-94

Appendix A: Alternative enhancement methods

This Appendix describes methods considered for the video enhancement method described in Section 6.3. These methods were based on approaches that had apparently performed well in other applications, but did not produce fast and reliable results as

expected. Here they are grouped according to techniques used, and reasons for rejection are given.

The methods did not all produce their results according to the criteria of Section 5.3, though each tried to reconstruct one-dimensional test data sampled in some irregular way, performing interpolation or averaging of input values as required without producing details not implied by the samples. All methods were iterative, and fast and reliable convergence were desired. One method, indicated in the description, was selected for its one-dimensional performance, then implemented for two-dimensional data.

A.1. Methods based on standard minimisation methods

A.1.1   *restore1dsd*

This method used the conjugate gradient method to minimise the sum of the mean square error in predicting the input pixel values, and a regularisation term. The latter was proportional to the sum of squared second differences.

The method was slow to interpolate, and oscillated when interpolation was attempted over large distances.

A.1.2   *restore1drj*

This method minimised the same quantity as *restore1dsd*, using the Jacobi iteration and a band matrix approximation for the matrix inverse, designed for strong regularisation.

The method was slow, and diverged when regularisation was weak. Halving the change on each iteration made convergence consistent but slowed it further.

A.1.3   *restore1dbp*

This method minimised the mean square error of sample prediction, using a back-projection algorithm extended to allow non-uniform sampling. (Back-projection is the use of the transpose of the matrix **A** of Equation 3 in deriving high-resolution pixel estimates from sample values. [Irani & Peleg 1991] gives an example of its use.) Regularisation was achieved by terminating the algorithm before convergence, on the assumption that the lower frequency components would converge more quickly.

The method sometimes converged rapidly, but not when blur was present in the sampling. It was poor at interpolation for undersampling, possibly because of a poor initial guess.

A.1.4 restore1drl

This method used the Richardson-Lucy algorithm [Richardson 1972], which is like back-projection but with multiplicative changes instead of additive. It behaved like *restore1dbp*, to which it is approximately equivalent when contrast is low.

A.1.5 imergesxnbp

This method was a 2-dimensional implementation of a back-projection algorithm, adjusted for non-uniform sampling as in [Irani & Peleg 1991]. The initial estimate was simply a zoomed copy of the ROI of the reference image. The iteration for solving the usual equation (3) was then

$$\mathbf{x'} = \mathbf{x} + \left\{ (\mathbf{A} \otimes \mathbf{A})^T (\mathbf{y} - \mathbf{A}\mathbf{x}) \right\} \oplus \left\{ c\mathbf{A}^T\mathbf{u} \right\} \qquad (A1)$$

where $\otimes$ and $\oplus$ are element-wise multiplication and division, $\mathbf{u}$ is a vector of ones and c is a relaxation constant for which higher values make convergence slower but more stable. Like *restore1dbp* it depended on termination for regularisation.

The method produced good results in a few iterations when sampling was uniform and no resolution increase was required. Higher resolutions required more iterations, while irregular sampling (likely when interlacing and motion are combined) produced artefacts or complete failure (division by zero for pixels missed by all samples). This method needs further investigation, to test the effect of introducing regularisation terms on its convergence.

A.2. Methods with extra smoothing steps

In these methods, the changes to the reconstructed values were chosen and scaled in each iteration according to one of the methods of Section A.1, but smoothed before the scaling was decided. In this way the results of early iterations were constrained to be smooth.

The smoothing was defined as convolution with the first L kernels in the sequence

$$(1/4, 2/4, 1/4)$$
$$(1/4, 0, 2/4, 0, 1/4)$$
$$(1/4, 0, 0, 0, 2/4, 0, 0, 0, 1/4)$$
$$\vdots$$

where the spacing of the non-zero entries in the Lth kernel is 2^{L-1}. (This choice is related to the wavelet methods in Sections 3.3 and 3.4.) L will be referred to as the level of smoothing. The overall impulse response of this smoothing is triangular and its width approximately doubles for successive levels. The strategy was usually to start with high-level smoothing and reduce the level to zero.

A.2.1   restore1djcgvs

This method was a variant of the conjugate gradient method (the one finally chosen) in which smoothing was enforced, and the level of smoothing was increased until convergence occurred, then set back to zero for the next attempt. It was tested with various limits for the number of iterations at a given smoothness level, and for the maximum smoothness level.

The method was generally much slower than the straight conjugate gradient method, perhaps because the conjugate gradient algorithm had to be restarted whenever the level of smoothness changed.

A.2.2   restore1dbpvs

This method used back-projection to minimise the sum of the mean square error in predicting the input pixel values, and a regularisation term. The latter was proportional to the sum of squared second differences. In the early iterations, the signal was restricted to be a combination of smoothed impulses, and the level of smoothness was reduced until there no restriction remained. It was expected that early restriction would allow more rapid convergence where there was sparse sampling.

The method was stable for a wide range of regularisation levels and handled interpolation well, but once zero smoothness was reached, the final stage suffered from slow convergence. This stage was needed to give the full resolution expected in well sampled parts of the signal. The problems with this approach are likely to affect many others when the density of samples varies over a signal or image.

Deciding the numbers of iterations to apply at different smoothness levels was a further problem with this method.

A.2.3   restore1dbpvs2

This method was a variant of *restore1dbpvs*, in which the level of smoothing was increased until convergence occurred, then set back to zero for the next attempt. It was tested with various limits for the number of iterations at a given smoothness level, and for the maximum smoothness level.

The method was sometimes faster but usually worse than *restore1dbpvs*.

A.2.4   restore1dsdvs

This method was similar to *restore1dbpvs*, but used a steepest-descent method instead of back-projection. It suffered from slow convergence.

A.2.5   restore1dsdvs2

This method was a variant of *restore1dsdvs*, in which the level of smoothing was increased until convergence occurred, then set back to zero for the next attempt. It was tested with various limits for the number of iterations at a given smoothness level, and for the maximum smoothness level.

The method was faster than *restore1dsdvs*, but often still slow.

A.2.6   restore1dsdvr

This method was a variant of *restore1dsdvs*, in which regularisation was stronger in early iterations. It showed faster convergence in those iterations but not when the regularisation was set to its final level.

A.2.7   restore1dsdar

This method was another variant of *restore1dsdvs* in which the regularisation term of the objective function was modified to make the diagonal of the matrix A in equation (3) more nearly constant. This tactic was expected to speed up convergence, but failed to do so when samples were sparse in some areas.

A.3.   Methods using wavelet analysis for filtering

In wavelet analysis of discrete-time signals [Daubechies 1992], a signal is represented as a combination of signals with different scales of detail. The separation of scales is done in such a way that coarser components can be represented by fewer samples without loss of information. The theory provides a way to restore the missing samples by interpolation as the components are recombined to recover the signal. It can be extended readily to two dimensions. The component signals can be further interpreted as combinations of functions of finite support, zero mean and various scales and positions (wavelets), so the approach is an alternative to Fourier analysis, sometimes with superior properties.

There are efficient algorithms for separating the components (analysis) and recombining them (synthesis). They each depend on processing one level of detail at a time. The basis step in analysis is to separate the signal into low-frequency and high-frequency parts, discarding half the samples in each. The high-frequency part constitutes the finest detail of the signal, while the low-frequency part is a coarser version of it. The low-frequency part is subject to the same process to get the next level of detail and an even coarser version of the signal. The process is repeated for as many levels as desired. The basic step is readily reversed, and can be applied at each level starting with the coarsest to perform synthesis.

There are many ways to design the filters in wavelet analysis and synthesis. Some lead to orthogonality of all individual wavelets, others to smoother wavelets, symmetric or anti-symmetric wavelets and so on. In this study wavelet analysis and synthesis were done using the "lifting" approach of [Sweldens 1995].

There is a variation of the wavelet technique ("undecimated" wavelets or "á trous" analysis) in which the number of samples is not reduced as the analysis proceeds, but the lengths of the filters are increased by moving the taps further apart. This variation removes the difficulty that the way a signal feature is distributed over the different wavelet scales depends strongly on its location, at the cost of increased data and computation at coarser scales.

A.3.1   restore1dwj

This method used the Jacobi method to minimise the sum of the mean square error in predicting the input pixel values, and a regularisation term. The latter was proportional to the sum of squared second differences.

The method differed from *restore1drj* by using a different approximation of the matrix, based on wavelet analysis. (See Section A.2.) Such an approach was used by [Lai & Vemuri 1997]. After wavelet analysis, each coefficient was multiplied by the intended frequency response of the inverse at the centre of the effective pass band of the corresponding wavelet. Then wavelet synthesis was applied.

The method was faster than *restore1dj* but convergence was less reliable, and did not interpolate sparse samples well. Modification of how the centre frequency was selected did not improve the method.

A.3.2   restore1datj

This method different from *restore1dwj* by using "á trous" wavelet analysis designed so that wavelet synthesis reduced to simple summation of the coefficient sequences. The analysis was based on linear interpolation.

The method converged well for moderate levels of regularisation, but was still poor at interpolation of sparse samples. Treatment of the start and end of the signal was suspected as a source of difficulties, but changes to this treatment (such as making the signal periodic) made little difference.

A.3.3   restore1datcj

This method was similar to *restore1datj* but used wavelet analysis based on cubic interpolation so that frequency bands were better separated. It converged slightly faster than *restore1datj* but was no improvement otherwise.

A.3.4   *restore1dw*

This method was similar to *restore1dwj* but used the pre-conditioned conjugate gradient method instead of Jacobian.

The method converged more often than *restore1dwj* and could interpolate sparse samples, but was very slow for larger gaps. When the number of levels in the wavelet analysis was varied, the best results were obtained for no levels at all (ie without wavelets).

A.3.5   *restore1dwc*

This method was like *restore1dw* but used wavelet analysis based on cubic interpolation so that frequency bands were better separated. It gave similar results to that method.

A.3.6   *restore1djcgds*

This method was a variant of the conjugate gradient method (the one finally chosen) in which smoothing was imposed by using partial wavelet analysis, and the amount of smoothing was increased until convergence occurred, then set back to none for the next attempt. It was tested with various limits for the number of iterations at a given smoothness level, and for the maximum smoothness level.

The method was generally much slower than the straight conjugate gradient method, perhaps because the conjugate gradient algorithm had to be restarted whenever the degree of smoothness changed.

A.3.7   *restore1dbpsm*

This method differed from *restore1dbp* by performing a smoothing step after each back-projection step. The smoothing was done via the wavelet coefficients, with wavelet analysis based on cubic interpolation.

Convergence was too slow.

A.4.   Methods using wavelet coefficients for regularisation

Wavelet analysis was described in Section A.3. The wavelet coefficients provide a measure of fine detail and regularisation can be based on them.

A.4.1   *restore1d*

This method minimised the sum of the mean square error in predicting the input pixel values, and a regularisation term. The latter was a weighted sum of squared wavelet coefficients, with highest weight given to the finest scale of detail and zero weight to all scales coarser than the last level found in analysis (a parameter of the method). The

wavelets were based on linear interpolation. The conjugate gradient method was used in the minimisation, and wavelet analysis calculations were encoded into it as matrix operations.

This method sometimes performed well, but it did not interpolate reliably where samples were sparse.

A.4.2 *restore1dc*

This method was similar to *restore1d* but used wavelets based on cubic interpolation so that frequency bands were better separated. Results were similar to those of *restore1d*.

A.4.3 *restore1dj*

This method was similar to *restore1d* but used the Jacobi iteration to minimise. It converged slowly and unreliably, diverging for weak regularisation. Changing the weighting of wavelet coefficients in regularisation made no difference to convergence.

A.4.4 *restore1dcj*

This method was similar to *restore1dj* but used wavelets based on cubic interpolation so that frequency bands were better separated. Results were similar to those of *restore1dj*.

A.5. Methods using projection onto convex sets (POCS)

The POCS approach is used when a solution is required that satisfies a given set of constraints without necessarily being optimal in any way. Each constraint is applied in turn, the present estimate being replaced by the nearest estimate that obeys the constraint. The approach is proven to converge to a correct solution (of which there may be more than one) if each constraint defines a convex set and the intersection of the sets is non-empty [Youla & Webb 1982]. The solution may or may not be reached in a finite number of iterations.

A.5.1 *restore1dp*

This method used the POCS method to force each error in predicting the input pixel values to be smaller than an error limit, and each second difference to be smaller than a roughness limit.

The method was very slow to converge, though sometimes results were fair reproductions of test data well before convergence. Interpolation of sparse samples was poor. The behaviour was sensitive to the ratio of the error and roughness limits. Setting the limits to zero in early iterations might be expected to speed up convergence in some cases but did not have this effect.

A.5.2   *restore1dpm*

This varied the method of *restore1dp* by applying one iteration for each error limit to one copy of the current estimate and one iteration for each roughness limit to another copy, then taking the mean of the two.

The method converged more slowly than *restore1dp* and did not interpolate sparse samples well.

A.5.3   *restore1dps*

This method was a variant of *restore1dp* in which constraints were enforced in blocks by surrogate projection. Surrogate projection finds the adjustment that would be made by projection for each constraint in a block, then applies a weighted mean of those adjustments, repeating for each block in a cycle [Yang & Murty 1992]. Parameters allowed relaxation and alteration of the sizes of the blocks.

The method did not produce any improvement over *restore1dp*. When the constraints were too tight for any solution, false convergence occurred.

A.5.4   *restore1dps2*

This method was a further variant of *restore1dps* in which each block of constraints was chosen to include prediction error and roughness limits from the same part of the output image.

Convergence was faster than for *restore1dp*, but still too slow. The results suggested that changes to output pixel values were affecting too many second differences around them, so the blocks of constraints were interacting too much.

DISTRIBUTION LIST

Techniques for Extracting High-resolution Still Images from Video Sequences

Robert Whatmough

AUSTRALIA

DEFENCE ORGANISATION

S&T Program
Chief Defence Scientist ⎫
FAS Science Policy ⎬ shared copy
AS Science Corporate Management ⎪
Director General Science Policy Development ⎭
Counsellor Defence Science, London (Doc Data Sheet)
Counsellor Defence Science, Washington (Doc Data Sheet)
Scientific Adviser to MRDC Thailand (Doc Data Sheet)
Scientific Adviser Joint
Navy Scientific Adviser
Scientific Adviser - Army (Doc Data Sheet and distribution list only)

Air Force Scientific Adviser
Director Trials

Aeronautical and Maritime Research Laboratory
Director
Mr. Trevor Adams, MOD
Mr. Stuart Sutherland, MOD

Electronics and Surveillance Research Laboratory
Director (Doc Data Sheet and distribution list only)

Chief of Surveillance Systems Division
Research Leader Imagery Systems
Head Image Analysis and Exploitation Group
Mr. Don Uksi
Dr. Robert Caprari
Ms Merrilyn Fiebig

Chief of Information Technology Division
Head C3 Information Systems Concepts Group, ITD
Dr. John Hildebrandt, ITD
Dr. Robert Prandolini, ITD

Chief of Electronic Warfare Division
Head EO Countermeasures Group
Dr. Brian Craig

Author(s): Mr. Robert Whatmough (6 copies)

DSTO Library and Archives
  Library Fishermans Bend (Doc Data Sheet)
  Library Maribyrnong (Doc Data Sheet)
  Library Edinburgh
  Australian Archives
  Library, MOD, HMAS Stirling
  Library, MOD, Pyrmont (Doc Data sheet only)

  US Defense Technical Information Center, 2 copies
  UK Defence Research Information Centre, 2 copies
  Canada Defence Scientific Information Service, 1 copy
  NZ Defence Information Centre, 1 copy
  National Library of Australia, 1 copy

Capability Systems Staff
  Director General Maritime Development (Doc Data Sheet only)
  Director General Land Development
  Director General Aerospace Development (Doc Data Sheet only)

Knowledge Staff
  Director General Command, Control, Communications and Computers (DGC4)
      (Doc Data Sheet only)

Navy
  COMNAVAUSUBGRPCSOCAPDEV, CMDR John Taubman
  DNC4ISREWSG, Attn: Mr Aiyel Chelliah

Army
  Stuart Schnaars, ABCA Standardisation Officer, Tobruk Barracks, Puckapunyal,
      3662 (4 copies)
  SO (Science), Deployable Joint Force Headquarters (DJFHQ) (L), MILPO Gallipoli
      Barracks, Enoggera QLD 4052 (Doc Data Sheet only)

Intelligence Program
  DGSTA Defence Intelligence Organisation
  Manager, Information Centre, Defence Intelligence Organisation
  Mr. T. O'Leary, Australian Imagery Organisation

Corporate Support Program
  Library Manager, DLS-Canberra

UNIVERSITIES AND COLLEGES
  Australian Defence Force Academy
      Library
      Head of Aerospace and Mechanical Engineering
      A/Prof Don Fraser
  Serials Section (M list), Deakin University Library, Geelong, 3217
  Hargrave Library, Monash University (Doc Data Sheet only)
  Librarian, Flinders University
  Prof. Mike Brooks, University of Adelaide
  Dr. Anton van den Hengel, University of Adelaide

OTHER ORGANISATIONS
  NASA (Canberra)
  AusInfo
  State Library of South Australia

OUTSIDE AUSTRALIA

ABSTRACTING AND INFORMATION ORGANISATIONS
  Library, Chemical Abstracts Reference Service
  Engineering Societies Library, US
  Materials Information, Cambridge Scientific Abstracts, US
  Documents Librarian, The Center for Research Libraries, US

INFORMATION EXCHANGE AGREEMENT PARTNERS
  Acquisitions Unit, Science Reference and Information Service, UK
  Library - Exchange Desk, National Institute of Standards and Technology, US

SPARES (5 copies)

Total number of copies: **69**

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA		1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)	
2. TITLE Techniques for Extracting High-resolution Still Images from Video Sequences		3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)	
4. AUTHOR(S) Robert Whatmough		5. CORPORATE AUTHOR Electronics and Surveillance Research Laboratory PO Box 1500 Edinburgh South Australia 5111 Australia	

6a. DSTO NUMBER DSTO-TR-1247	6b. AR NUMBER AR-012-089	6c. TYPE OF REPORT Technical Report	7. DOCUMENT DATE December 2001
8. FILE NUMBER B9505-21-224	9. TASK NUMBER JNT 99/016	10. TASK SPONSOR Refer Task Manager	11. NO. OF PAGES 41	12. NO. OF REFERENCES 50

13. URL on the World Wide Web http://www.dsto.defence.gov.au/corporate/reports/DSTO-TR-1247.pdf	14. RELEASE AUTHORITY Chief, Surveillance Systems Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, SALISBURY, SA 5108

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CASUAL ANNOUNCEMENT	Yes

18. DEFTEST DESCRIPTORS

Video frames, Image processing, High resolution

19. ABSTRACT

This Report describes techniques for extracting a single frame or part of a frame from a video image sequence and combining information from other frames to enhance the resolution of the result. The result depends on accurate alignment of the frames and construction of a detailed image consistent with the coarse data in the frames. Techniques for the alignment and the construction steps, and theoretical limitations on the final resolution, are considered.